VOLUME 22 · NUMBER 12 · JUNE 15 2004

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Determining the Relationship Between Toxicity and Quality of Life in an Ovarian Cancer Chemotherapy Clinical Trial

Lorna Butler, Monica Bacon, Mark Carey, Benny Zee, Dongsheng Tu, and Andrea Bezjak

From the Dalhousie University and Queen Elizabeth II Health Sciences Centre, Halifax, Nova Scotia; National Cancer Institute of Canada Clinical Trials Group, Kingston; Department of Obstetrics/Gynecology, University of Western Ontario, London; and Toronto Hospital Network Princess Margaret Hospital and University of Toronto, Toronto, Ontario, Canada.

Submitted January 17, 2003; accepted March 18, 2004.

Supported by the Socio-Behavioral Research Network of the Canadian Cancer Society and National Cancer Institute of Canada.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Lorna Butler, PhD, Room 122, Forrest Building, Dalhousie University, Halifax, Nova Scotia B3L 4H2, Canada; e-mail: lorna.butler@dal.ca.

0732-183X/04/2212-2461/$20.00

DOI: 10.1200/JCO.2004.01.106

ABSTRACT

Purpose
This analysis of data from a randomized trial of chemotherapy in epithelial ovarian cancer sought to determine whether a relationship exists between the presence and severity of the most commonly observed toxic effects and the corresponding quality of life (QOL) items.

Patients and Methods
One hundred fifty-two eligible patients accrued from Canada by the National Cancer Institute of Canada Clinical Trials Group on a randomized trial of paclitaxel and cisplatin versus cyclophosphamide/cisplatin were included in the analysis. Toxicity to the chemotherapeutic treatments was subjectively evaluated using a trial-specific checklist for ovarian cancer and the European Organization for Research and Treatment of Cancer QLQ C30+3 questionnaire. Assessments were conducted at baseline, before each cycle of treatment (3 weeks), and at each 3-month follow-up during the next 2 years (or until progression).

Results
The most frequently observed symptoms experienced during or shortly following chemotherapy were neurosensory loss, lethargy, nausea, vomiting, and alopecia. Regression analyses revealed that change scores of QOL items related to motor weakness and gastrointestinal pain were common predictors for the change global QOL score during protocol treatment; and change scores of QOL items related to lethargy or fatigue and change toxicity grade of mood predicted the change global QOL score after patients were off treatment.

Conclusion
The use of the European Organization for Research and Treatment of Cancer QLQ C30+3 and trial-specific checklist was able to assess the effect of expected toxicities on patient's QOL during and following treatment, and so may be useful in addressing the concerns regarding methodological issues that have limited the acquisition of prospective, longitudinal treatment-related toxicity data.

J Clin Oncol 22:2461-2468. © 2004 by American Society of Clinical Oncology

INTRODUCTION

In 2002, approximately 2,500 Canadian women were diagnosed with ovarian cancer, and approximately 1,500 died from this disease.[1] These and other women who are diagnosed with ovarian cancer experience changing symptomatology, both during the course of their treatment and for the remainder of their lives as the disease progresses.

Measurement of symptom experience is not a straightforward process. The validity and reliability of the information-gathering method and the informant used for rating toxicity has engendered considerable debate.[2-4] One complicating factor is that data managers often have varying perspectives on toxic effects. The result is that intrarater reliability can be high, yet interrater reliability can be quite variable.[2] Recently, there has been an increasing interest in obtaining this information directly from the individual cancer patient, rather than relying on observers. Considerable interest has arisen in how to determine the impact of treatment effects using patient ratings of their own functional status through the measurement of patients' perceived quality

of life (QOL). A patient's perceptions of the toxicity of a treatment or other symptoms experienced during therapy as measured by QOL may, however, differ considerably from the assessment by health care providers.

Another important problem in clinical practice is to understand how meaningful toxic effects, whether assessed by clinicians or patients, are to patients' overall QOL. At the present time, clinicians struggle to provide accurate information to patients as to how the effects of treatments that are designed to lessen symptoms or even lead to prolongation of survival will affect their QOL. This highlights the importance of clinicians being able to interpret and use QOL data in a meaningful way in their practice.[5] Moreover, to what extent the patient's experience of functional changes is due to treatment toxicity and how these changes affect their overall QOL has received little consideration from an outcomes-management perspective.

In this article, we report results of a joint research project carried out by researchers from the Canadian Sociobehavioural Research Network and the National Cancer Institute of Canada Clinical Trials Group (NCIC-CTG). In this project, using data collected from the European-Canadian Intergroup trial comparing combination paclitaxel/cisplatin versus combination cyclophophamide/cisplatin chemotherapy in women with advanced epithelial ovarian cancer (OV10), we sought to assess the agreement between the symptoms or toxic effects (as recorded on case report forms [CRFs]) and QOL items that were associated with the most frequently observed symptoms. Further, these data were explored to determine whether there was an association between the toxicity type and the corresponding items in the QOL scale with the patients' change in their global QOL, with the aim of identifying those symptoms that were most predictive of change in global QOL.

PATIENTS AND METHODS

Trial Details and End Point Evaluations

In total, the OV10 study randomized 680 patients with histologically confirmed International Federation of Gynecology and Obstetrics (FIGO) stage IIB, IIC, III, or IV epithelial ovarian cancer.[6] Among them, 152 eligible patients were from Canada and were recruited by NCIC-CTG. Informed consent was obtained according to local human ethics review committee requirements. Toxicity and QOL data were collected prospectively on all patients. The European-Canadian "Intergroup Phase III Comparison of a Combination of Taxol/Platinum and a Combination of Cyclophophamide/Platinum Chemotherapy in the Treatment of Advanced Epithelial Ovarian Cancer" (OV10) recruited 680 patients with total abdominal hysterectomy and bilateral salpingo-oophorectomy, and histologically confirmed F1GO stage IIB, IIC, III, or IV epithelial ovarian cancer.

Following the medical review of the patient at scheduled intervals, patient symptoms were recorded on CRFs by research staff utilizing the NCIC-CTG Expanded Common Toxicity Criteria (CTC),[7] providing more than 100 available symptoms. Each item is graded using a 0 to 4 scale, with grade 0 meaning absence of toxicity, and grade 4 meaning life-threatening or disabling. The information (worst grade during the given period of assessment and relation to study agents) on a variety of symptoms was explicitly requested on the CRF during the treatment period at the end of each cycle for a maximum of nine cycles. Any additional toxic effects, beyond those explicitly solicited, could also be recorded on the form. The symptoms listed on the CRF were (in alphabetical order): anorexia, allergy, alopecia, arthralgia, asymptomatic bradycardia, bronchospasm, constipation, cystitis, diarrhea, dyspnea, edema, fatigue, fever, flushing, hearing loss or tinnitus, hemorrhage, hypotension, infection, local toxicity at the intravenous catheter site, motor impairment, mucositis, myalgia, nausea, pain, partial bowel obstruction, rigor, sensory impairment, skin rash, urticaria, and vomiting. A somewhat shorter list of symptoms was included on the baseline CRF, and once again, any additional symptoms not noted in the list could be added. Postchemotherapy symptoms were also captured on the CRF every 3 months for 2 years after patients were off treatment, or until progression.

QOL was subjectively evaluated at the same time points using the European Organization for Research and Treatment of Cancer (EORTC) QLQ C30+3 and a trial-specific checklist for ovarian cancer. The core questionnaire (QLQ C30+3) has five functional dimensions (physical, role, cognitive, emotional, and social), three symptom dimensions (fatigue, pain, and nausea/vomiting), a global QOL scale, and six single items concerning appetite loss, constipation, diarrhea, dyspnea, sleep disturbance, and financial consequence of the disease and treatment. EORTC-QLQ C30 is a well-established questionnaire that has been psychometrically validated in patients with numerous cancer types including ovarian cancer.[8] Three additional items (questions 31, 32, and 33) of a developmental nature were added to measure, respectively, the overall physical condition, overall health, and overall QOL. The trial-specific checklist was designed for this study and consists of a series of 11 questions that provided additional details on symptom-related distress. Five additional questions designed and added as a subjective significance module to elicit patients' opinions about the perceived physical discomfort and overall QOL were not included in this analysis. Most of the questions use a four-point Likert scale, with options ranging from "not at all" (1) to "very much" (4), except for questions 1 through 7, which use a two-point scale with "no" or "yes" answers and questions 31 to 33, which use a seven-point scale with options ranging from "very poor" to "excellent." Questions were asked in reference to the past week (items 1 through 7) or during the past week (items 8 through 44).

Statistical Analysis

The treatment-related incidences of toxicities on the CRF during protocol therapy were assessed to identify the most frequently observed symptoms, which were defined as those toxicities which had 10% or higher incidences during the protocol treatment (or the follow-up assessments) in either arm (Table 1). These most frequently observed symptoms were reexamined to identify corresponding QOL questions matched with these toxicities so comparisons between them could be made. In addition, two frequently observed QOL items were added—urinary frequency and incontinence. A total of 18 pairs of toxicities and QOL questions were identified (Table 2), which included the multiple matches between mood toxicity and four QOL questions (questions 21 to 24). The single questions in the QOL questionnaire and checklist (rather than domains) were used because they were more

Table 1. Observed Incidence of Treatment-Related Toxicity

Toxicity	Cisplatin and Cylophosphamide (assessable patients, n = 73)		Cisplatin and Taxol (assessable patients, n = 79)	
	Frequency	%	Frequency	%
During protocol treatment				
Cardiovascular, edema	5	7	14	18
Flu-like symptoms				
Arthralgia	3	4	33	42
Lethargy	64	88	70	89
Myalgia	13	18	42	53
Gastrointestinal				
Anorexia	38	52	48	61
Diarrhea	20	27	27	34
Heartburn	8	11	7	9
Nausea	69	95	71	90
Pain	7	10	12	15
Stomatitis	16	22	17	22
Taste altered	11	15	22	28
Vomiting	63	86	56	71
Neurologic				
Constipation	36	49	58	73
Cortical			1	1
Dizziness	10	14	11	14
Headache	15	21	12	15
Hearing	23	32	27	34
Insomnia	1	1	8	10
Mood	11	15	10	13
Motor	15	21	23	29
Sensory	34	47	70	89
Vision	4	5	9	11
Skin				
Alopecia	65	89	75	95
Facial flushing	3	4	25	32
Rash/itch	0	0	13	16
During patient follow-up				
Flu-like symptom (lethargy)	9	12	12	15
Neurologic				
Hearing	10	14	5	6
Sensory	30	41	62	78
Skin (alopecia)	5	7	21	27

relevant to the individual symptom assessments. Data at cycles three and six during the protocol treatment, and at 3 and 6 months after patients were off treatment, were used to assess the agreement between the identified symptom from the CRF and corresponding QOL question, since most acute and delayed symptoms could be found at these time points with minimum amount of missing data.

A preliminary analysis of the data showed that there were very few patients who had grade 2 or higher toxic effects or who reported a raw score 3 or higher in their QOL assessment, especially at baseline. Therefore, a single κ statistic was not appropriate in describing the relationship between toxicity and QOL assessments.[6] Following the suggestions of Feinstein and Cicchette,[9,10] both the observed toxicity and QOL measures were first divided into two categories: mild or none, and severe or moderate. A toxicity was classified as "mild or none" if the grade was ≤ 1 or the QOL raw score was 1 or 2, and as "severe or moderate" if the toxicity grade was ≥ 2 or the QOL raw score was 3 or 4. The index of the agreement between the toxicity and corresponding QOL question was then assessed separately for these two categories (termed respectively as the degree of agreement in "severe or moderate" or "mild or none" category). Basically, the degree of agreement in the mild or none category was calculated as two times the number of patients with both toxicity and QOL assessments classified into the mild or none category divided by the numbers of patients classified as mild or none based on toxicity only, plus the number of patients classified as mild or none based on the QOL assessment only. The degree of agreement in the severe or moderate category was calculated similarly. The 95% CIs for the degree of agreements in either the severe or moderate or mild or none category were calculated using the method and program provided by MacKinnon.[11]

To determine whether the symptoms investigated and their corresponding questions in the QOL questionnaire were also significantly associated with patients' change in global QOL, stepwise regression models were used to determine which toxicities and QOL questions were significantly associated both with baseline global QOL and the change from the baseline in the global QOL. As noted earlier, questions 31 to 33 measuring global QOL were developmental in nature, and question 31 has since been removed from the instrument. Thus, all the regression analyses were performed separately with rescaled scores (to 0 through 100) based on questions 32 and 33 only, or including all three of the questions,[31-33] as the dependent variable. A symptom or QOL question was entered into the model and remained in the model when it was significant at the 0.1 level, while controlling for other symptoms or QOL questions already in the model, until no further factor was significant. The amount of variation explained by the factors in the model was measured by the coefficient of determination (R^2).

RESULTS

Agreement Between Pairs of Matched CRF Symptoms and QOL Questions

At baseline, before the initiation of the treatment, the analysis revealed a close agreement in the mild or none category between the symptoms recorded on the CRF and the paired QOL questions, with greatest degree of agreement ranging between 0.80 (95% CI, 0.75 to 0.86) to 0.98 (95% CI, 0.92 to 0.99), as presented in Table 3. Two of the pairs, lethargy with QOL question 18 (degree = 0.72) and mood with QOL question 22 (degree = 0.73) were slightly weaker in agreement than all other pairs. There were few severe or moderate symptoms or QOL assessments reported at baseline, with only two pairs demonstrating a moderate agreement (degree of agreement higher than 0.40) in the severe or moderate category. Those pairs were: constipation, with QOL question 16 (degree = 0.44), and lethargy, with QOL question 18 (degree = 0.44).

During protocol treatment, at the end of both cycles 3 and 6, all but one of the symptom and QOL pairs demonstrated marked agreement in the mild or none category, with the degree of agreement ranging between 0.71 and

Table 2. Pairs of Most Frequently Observed Toxicities and Matching QOL Questions

Toxicity	Grading	QOL Question No.	Description
Neurosensory (NE SEN)	1-2, mild to moderate paresthesias ≥ 3, sensory loss or paresthesias that interferes with function	41	Have changes in sensation in your fingers or toes been a problem?
Gastrointestinal pain (GI PAI)	1-2, mild to moderate ≥ 3, severe	40	Have you bothered by stomach cramps?
Alopecia (SK ALO)	1-2, mild to pronounced- or total-head hair loss ≥ 3, total body hair loss	42	Have you been bothered by hair loss?
Urinary frequency (GU FRE)	1-2, frequency of urination or nocturia < hourly ≥ 3, frequency with urgency and nocturia ≥ hourly	38	Have you been bothered by frequent urination?
Incontinence (GU INC)	1-2, mild to moderate ≥ 3, severe	39	Have you had a loss of bladder control?
Myalgia (FL MYA)	1-2, mild to moderate ≥ 3, severe	35	Have aches or pains in your muscles been a problem?
Constipation (NE CON)	1-2, mild to moderate ≥ 3, severe constipation to ileus > 96 hours	16	Have you been constipated?
Motor (NE MOT)	1-2, subjective weakness to mild objective weakness ≥ 3, objective weakness with impairment of function	12	Have you felt weak?
Diarrhea (GI DIA)	1-2, increase of 2-6 stools/day ≥ 3, increase of ≥ 7 stools/day or grossly bloody stool	17	Have you had diarrhea?
Vomiting (GI VOM)	1-2, 1-5 episodes in 24 hours ≥ 3, ≥ 6 episodes/24 hours or require parenteral support	15	Have you vomited?
Nausea (GI NAU)	1-2, reasonable to significantly decreased intake but can eat ≥ 3, no significant intake	14	Have you felt nauseated?
Anorexia (GI ANO)	1-2, mild to moderate ≥ 3, severe dehydration	13	Have you lacked appetite?
Lethargy (FL LET)	1-2, mild to moderate or fall of 1-2 levels in performance status ≥ 3, severe or fall of 3 levels in performance status	18	Were you tired?
Mood (NE MOO)	1-2, mild to moderate anxiety or depression ≥ 3, severe anxiety or depression to suicidal ideation	21-24	(21) Did you feel tense? (22) Did you worry? (23) Did you feel irritable? (24) Did you feel depressed?
Insomnia (NE INS)	1-2, mild to moderate ≥ 3, severe	11	Have you had trouble sleeping?

Abbreviation: QOL, quality of life.

0.93. The pair assessing hair loss (symptom alopecia and QOL question 42) was the one exception. At cycle 3, the degree of agreement for this pair was 0.50, and at cycle 6, 0.37. The agreement in the severe or moderate category was very weak for all pairs (almost all degrees of agreement were less than 0.5; Table 3).

At 3 and 6 months' follow-up, after patients were off front-line chemotherapy, a strong agreement (degree of agreement higher than 0.8) in the mild or none category existed between most of the pairs, with maximum degree of agreement at 0.88. Three of the pairs had slightly weaker degrees of agreement, including the pairs related to neurosensory symptoms and QOL question 41 (3-month follow-up degree = 0.54; 6-month follow-up degree = 0.72), myalgia and QOL question 35 (3-month follow-up degree = 0.74; 6-month follow-up degree = 0.79), and lethargy and QOL question 18 (3-month follow-up degree = 0.76; 6-month follow-up degree = 0.79). The degree of agreement between alopecia and QOL question 42 was 0.78 at 3 months' follow-up. Agreement was poor with most pairs in the severe or moderate category at both the month-3 and month-6 follow-up time points, except for the pair of neurosensory and QOL question 41, which had a degree of agreement of 0.62 at 3 months' follow-up and 0.57 at 6 months' follow-up (Table 3).

Table 3 also presents when patients scored a symptom as severe or moderate using the QOL definition; this was less often graded as severe or moderate (grade 2 or more) by the research staff when the same symptom was recorded on the CRF.

Predicting Baseline Global QOL Scores Based on Baseline Grades of the Most Frequently Observed Toxicities and Scores of Corresponding QOL Questions

When the baseline rescaled score for the global QOL as measured by all three questions (questions 31 to 33) was used as the dependent variable, and baseline grades of the most frequently observed symptoms and the baseline raw scores of the corresponding QOL questions were entered as independent variables in the stepwise regression analysis,

Table 3. Agreement Between Toxicity and QOL Assessments

Pair and Assessment Time	MNTOX and MNQOL		MNTOX and SMQOL		SMTOX and MNQOL		SMTOX and SMQOL		MN		SM	
	No.	%	No.	%	No.	%	No.	%	d	95% CI	d	95% CI
Sensory and QOL Q 41												
Baseline	146	94.1	6	3.9	0	0	0	0	0.98	0.92-0.99	0	0-0
Cycle 3	116	80.0	26	17.9	2	1.4	1	0.7	0.89	0.85-0.93	0.07	0-0.19
Cycle 6	80	63.5	21	16.7	7	5.6	18	14.3	0.85	0.80-0.91	0.56	0.42-0.71
Month 3	28	24.6	43	37.7	4	3.5	39	34.2	0.54	0.43-0.66	0.62	0.52-0.72
Month 6	41	43.6	24	25.5	8	8.5	21	22.3	0.72	0.63-0.81	0.57	0.43-0.70
Gastrointestinal pain and QOL Q 40												
Baseline	107	70.4	28	18.4	7	4.6	10	6.6	0.86	0.81-0.90	0.36	0.20-0.53
Cycle 3	119	82.1	19	13.1	3	2.1	4	2.8	0.91	0.88-0.95	0.27	0.06-0.48
Cycle 6	106	84.1	20	15.9	0	0	0	0	0.91	0.88-0.95	0	0-0
Month 3	81	71.1	29	25.4	2	1.8	2	1.8	0.84	0.78-0.89	0.11	0-0.26
Month 6	69	73.4	21	22.3	2	2.1	2	2.1	0.86	0.80-0.91	0.15	0-0.33
Alopecia and QOL Q 42												
Baseline	144	94.7	8	5.3	0	0	0	0	0.97	0.95-0.99	0	0-0
Cycle 3	35	24.1	36	24.8	35	24.1	39	26.9	0.50	0.40-0.60	0.52	0.43-0.62
Cycle 6	23	18.3	23	18.3	55	43.7	25	19.8	0.37	0.26-0.48	0.39	0.28-0.50
Month 3	71	68.3	33	29.0	8	7.0	2	1.8	0.78	0.71-0.84	0.09	0-0.20
Month 6	71	75.5	21	22.3	1	1.1	1	1.1	0.87	0.81-0.92	0.08	0-0.24
Urinary frequency and QOL Q 38												
Baseline	127	83.6	24	15.8	0	0	1	0.7	0.91	0.88-0.95	0.08	0-0.22
Cycle 3	119	82.1	24	16.6	1	0.7	1	0.7	0.91	0.87-0.94	0.02	0-0.21
Cycle 6	107	84.9	19	15.1	1	0.9	0	0	0.92	0.88-0.96	0	0-0
Month 3	86	75.4	27	23.7	0	0	0	0	0.86	0.81-0.91	0	0-0
Month 6	74	78.7	20	21.3	0	0	0	0	0.88	0.83-0.93	0	0-0
Incontinence and QOL Q 39												
Baseline	143	94.1	9	5.9	0	0	0	0	0.97	0.95-0.99	0	0-0
Cycle 3	125	86.2	20	13.8	0	0	0	0	0.93	0.89-0.96	0	0-0
Cycle 6	109	86.5	17	13.5	0	0	0	0	0.93	0.89-0.96	0	0-0
Month 3	86	75.4	28	24.6	0	0	0	0	0.86	0.81-0.91	0	0-0
Month 6	73	77.7	21	22.3	0	0	0	0	0.87	0.82-0.93	0	0-0
Myalgia and QOL Q 35												
Baseline	129	84.9	21	13.8	0	0	2	1.3	0.92	0.89-0.96	0.16	0-0.35
Cycle 3	103	71.0	29	20.0	8	5.5	5	3.5	0.85	0.80-0.90	0.21	0.06-0.37
Cycle 6	97	77.0	17	13.5	8	6.4	4	3.2	0.89	0.84-0.93	0.24	0.05-0.44
Month 3	66	57.9	46	40.4	0	0	2	1.8	0.74	0.67-0.81	0.08	0-0.18
Month 6	61	64.9	31	33.0	0	0	2	2.1	0.79	0.73-0.87	0.11	0-0.26
Constipation and QOL Q 16												
Baseline	106	69.7	29	19.1	4	2.6	13	8.6	0.87	0.82-0.91	0.44	0.28-0.60
Cycle 3	92	63.5	23	15.9	19	13.1	11	7.6	0.81	0.76-0.87	0.34	0.19-0.49
Cycle 6	86	68.3	28	14.3	18	14.3	4	3.2	0.83	0.77-0.88	0.18	0.03-0.34
Month 3	79	69.3	34	29.8	1	0.9	0	0	0.82	0.76-0.88	0	0-0
Month 6	62	66.0	27	28.7	2	2.1	3	3.2	0.81	0.74-0.88	0.17	0-0.34
Motor weakness and QOL Q 12												
Baseline	102	67.1	50	32.9	0	0	0	0	0.80	0.75-0.86	0	0-0
Cycle 3	100	69.0	39	25.5	4	2.8	4	2.8	0.83	0.78-0.88	0.16	0.02-0.30
Cycle 6	88	69.8	34	25.4	3	2.4	3	2.4	0.83	0.78-0.89	0.15	0-0.29
Month 3	76	66.7	36	30.7	2	1.8	1	0.9	0.80	0.74-0.89	0.05	0-0.15
Month 6	67	71.3	26	27.7	1	1.1	0	0	0.83	0.77-0.89	0	0-0
Diarrhea and QOL Q 17												
Baseline	131	86.2	17	11.2	1	0.7	3	2.0	0.94	0.91-0.97	0.25	0.02-0.48
Cycle 3	120	82.8	19	12.1	1	0.7	5	3.5	0.92	0.89-0.96	0.33	0.12-0.55
Cycle 6	109	86.5	16	12.7	1	0.8	0	0	0.93	0.89-0.96	0	0-0
Month 3	84	73.7	27	23.7	1	0.9	2	1.8	0.86	0.81-0.91	0.13	0-0.28
Month 6	74	78.7	20	21.3	0	0	0	0	0.88	0.83-0.93	0	0-0
Vomiting and QOL Q 15												
Baseline	139	91.5	10	6.7	1	0.7	2	1.3	0.96	0.94-0.98	0.27	0-0.56
Cycle 3	102	70.3	10	6.9	22	15.2	11	7.6	0.86	0.82-0.91	0.41	0.24-0.57
Cycle 6	81	64.3	18	14.3	20	15.9	7	5.6	0.81	0.75-0.87	0.27	0.11-0.43
Month 3	87	76.3	24	21.1	1	0.9	2	1.8	0.87	0.83-0.92	0.14	0-0.31
Month 6	74	78.7	18	19.2	0	0	2	2.1	0.87	0.84-0.94	0.18	0-0.40

(continued on following page)

Table 3. Agreement Between Toxicity and QOL Assessments (continued)

Pair and Assessment Time	MNTOX and MNQOL		MNTOX and SMQOL		SMTOX and MNQOL		SMTOX and SMQOL		MN		SM	
	No.	%	No.	%	No.	%	No.	%	d	95% CI	d	95% CI
Nausea and QOL Q 14												
Baseline	128	84.2	17	11.2	4	2.6	3	2.0	0.92	0.89-0.96	0.22	0.01-0.43
Cycle 3	87	60.1	15	10.3	30	20.7	13	9.0	0.79	0.74-0.85	0.37	0.22-0.51
Cycle 6	62	49.2	15	11.9	36	28.6	13	10.3	0.71	0.63-0.79	0.34	0.20-0.47
Month 3	87	76.3	25	21.9	0	0	2	1.8	0.87	0.83-0.92	0.07	0-0.21
Month 6	72	76.6	20	21.3	0	0	2	2.1	0.88	0.83-0.93	0.17	0-0.37
Anorexia and QOL Q 13												
Baseline	101	66.5	42	27.6	1	0.7	8	5.3	0.82	0.77-0.88	0.27	0.12-0.42
Cycle 3	108	74.5	22	15.2	8	5.5	7	4.8	0.88	0.83-0.92	0.32	0.14-0.50
Cycle 6	94	74.6	20	15.9	9	7.1	3	2.4	0.87	0.82-0.91	0.17	0-0.34
Month 3	84	73.7	29	25.4	0	0	1	0.9	0.85	0.80-0.91	0.06	0-0.18
Month 6	70	74.5	23	24.5	0	0	1	1.1	0.86	0.80-0.92	0.08	0-0.23
Lethargy and QOL Q 18												
Baseline	73	48.0	47	30.9	10	6.6	22	14.5	0.72	0.65-0.79	0.44	0.31-0.56
Cycle 3	80	55.2	28	19.3	19	13.1	18	12.4	0.77	0.71-0.84	0.43	0.30-0.57
Cycle 6	76	60.3	26	20.6	14	11.1	10	7.9	0.79	0.73-0.85	0.33	0.18-0.49
Month 3	66	57.9	40	35.1	1	0.9	7	6.1	0.76	0.69-0.83	0.25	0.10-0.41
Month 6	59	62.8	30	31.9	2	2.1	3	3.2	0.79	0.71-0.86	0.16	0-0.32
Mood and QOL Q 21												
Baseline	101	66.5	40	26.3	6	4.0	5	3.3	0.81	0.76-0.87	0.18	0.04-0.31
Cycle 3	111	76.6	24	16.6	6	3.5	5	3.5	0.88	0.84-0.93	0.26	0.08-0.44
Cycle 6	95	75.4	23	18.3	6	4.8	2	1.6	0.87	0.82-0.94	0.12	0-0.27
Month 3	74	64.9	36	31.6	2	1.8	2	1.8	0.80	0.73-0.86	0.10	0-0.22
Month 6	65	69.2	26	27.7	0	0	3	3.2	0.83	0.77-0.90	0.19	0-0.37
Mood and QOL Q 22												
Baseline	84	55.3	57	37.5	5	3.3	6	4.0	0.73	0.67-0.80	0.16	0.05-0.28
Cycle 3	108	74.5	27	18.6	5	3.5	5	3.5	0.87	0.83-0.92	0.24	0.07-0.41
Cycle 6	95	75.4	23	18.3	7	5.6	1	0.8	0.86	0.82-0.91	0.06	0-0.18
Month 3	75	65.8	35	30.7	2	1.8	2	1.8	0.80	0.74-0.86	0.10	0-0.22
Month 6	66	70.2	25	26.6	0	0	3	3.2	0.84	0.78-0.90	0.19	0-0.38
Mood and QOL Q 23												
Baseline	120	80.0	21	13.8	9	5.9	2	1.3	0.89	0.58-0.93	0.12	0-0.27
Cycle 3	117	80.7	18	12.4	5	3.5	5	3.5	0.91	0.87-0.95	0.30	0.10-0.50
Cycle 6	98	77.8	20	15.9	8	6.4	0	0	0.88	0.83-0.92	0	0-0
Month 3	80	70.2	30	26.3	4	3.5	0	0	0.82	0.77-0.88	0	0-0
Month 6	72	76.6	19	20.2	0	0	3	3.2	0.88	0.83-0.94	0.24	0.02-0.46
Mood and QOL Q 24												
Baseline	112	73.7	29	19.1	7	4.6	4	2.6	0.86	0.82-0.91	0.18	0.03-0.34
Cycle 3	112	77.2	22	15.9	6	4.1	4	2.8	0.89	0.85-0.93	0.22	0.04-0.40
Cycle 6	99	78.6	19	15.1	8	6.4	0	0	0.88	0.84-0.92	0	0-0
Month 3	81	71.1	29	25.4	2	1.8	2	1.8	0.84	0.78-0.90	0.11	0-0.26
Month 6	72	76.6	19	20.2	0	0	3	3.2	0.88	0.83-0.94	0.24	0.02-0.46
Insomnia and QOL Q 11												
Baseline	108	71.1	43	28.3	0	0	1	0.7	0.83	0.79-0.88	0.04	0-0.13
Cycle 3	102	70.3	41	28.3	1	0.7	1	0.7	0.83	0.78-0.88	0.04	0-0.13
Cycle 6	89	70.6	34	27.0	1	0.8	2	1.6	0.84	0.78-0.89	0.07	0-0.23
Month 3	79	69.3	34	29.8	0	0	1	0.9	0.82	0.76-0.88	0.05	0-0.16
Month 6	64	68.1	30	31.9	0	0	0	0	0.81	0.74-0.88	0	0-0

Abbreviations: QOL, quality of life; MNTOX, "mild or none" toxicity (grade 0/1); MNQOL, "mild or none" QOL assessment (raw score 1/2); SMQOL, "severe or moderate" toxicity (grade ≥ 2); SMTOX, "severe or moderate" toxicity (grade ≥ 2); MN, "mild or none" category; SM, "severe or moderate" category; d, degree of agreement; Q, question number.

the final regression model retained only four QOL questions: questions 12 (related to motor weakness), 13 (related to anorexia), 24 (related to mood), and 40 (related to gastrointestinal pain), which explained 58% of the variance in the baseline global QOL ($F_{4134} = 46.85$; $P < .0001$). When only QOL questions 32 and 33 were used as the measurement of global QOL, the same four questions plus question 15 (related to vomiting) were in the final model. These five

variables explained 60% of the variance in baseline global QOL ($F_{9130} = 39.74$; $P < .0001$).

Association Between Change Scores From Baseline of the Global QOL and Change Grades From Baseline of the Most Frequently Observed Toxicities and Change Scores From Baseline of the Corresponding QOL Questions

At cycle 3 during the protocol treatment, the change QOL score from baseline of questions 12 (related to motor weakness), 35 (related to myalgia), 40 (related to gastrointestinal pain), and 42 (related to alopecia), and the change toxicity grade of urinary frequency were found predictive of the change global QOL score from baseline regardless of whether questions 31 to 33 or questions 32 and 33 only were used as the measurement of the global QOL. The change toxicity grade of insomnia was another predictive variable when the global QOL was measured by questions 31 to 33, while the analysis with questions 32 and 33 as the measurement of global QOL retained two additional variables: the change toxicity grade of constipation and myalgia. The R^2 of the final models in both cases were relatively the same ($R^2 = 62\%$ when questions 31 to 33 were used, and $R^2 = 65\%$ when questions 32 and 33 were used).

At cycle 6 during the protocol treatment, the QOL questions 12 (related to motor weakness), 13 (related to anorexia), 17 (related to diarrhea), 18 (related to lethargy or fatigue), and 40 (related to gastrointestinal pain) were in the final model when QOL questions 31 to 33 were used to measure the global QOL ($R^2 = 60\%$). When QOL questions 32 and 33 alone were the measurement of the global QOL, the final model included the following: QOL questions 12 (related to the motor weakness), 13 (related to anorexia), 17 (related to diarrhea), 18 (related to lethargy or fatigue), 24 (related to mood), and 38 (related to urinary frequency), and toxicity for vomiting ($R^2 = 65\%$).

After the patients were off treatment, at month-3 follow-up, three toxicity items (gastrointestinal pain, motor weakness, and mood) and six QOL questions (12 related to motor weakness, 17 related to diarrhea, 18 related to lethargy or fatigue, 24 related to mood, 39 related to incontinence, and 41 related to neurosensory symptoms) were in the final model when QOL questions 31 through 33 were the measurement of the global QOL ($R^2 = 72\%$). Almost all of the same variables were retained in the final model when the QOL questions 32 and 33 were used to measure the global QOL, except QOL question 41 (related to neurosensory symptoms) was replaced by QOL questions 16 (related to constipation) and 42 (related to alopecia: $R^2 = 73\%$). At 6 months' follow-up, the variables retained in the final models when either of the methods was used to measure the global QOL included the toxicity of mood and QOL questions 18 related to lethargy, and 41 related to neurosensory symptoms. Additional variables were toxicities of insomnia, vomiting, and lethargy, and QOL question 40 related to gastrointestinal pain when the QOL questions 31 through 33 were used to measure of the global QOL ($R^2 = 66\%$) and nausea, and QOL question 23 related to mood when the QOL questions 32 and 33 were the measurement of the global QOL ($R^2 = 60\%$).

DISCUSSION

Ovarian cancer can be a devastating condition for a woman, not only because of the physical and symptomatic toll the disease can exert, but also because of the physical and emotional upheaval and toxic effects caused by the treatments. Until recently, these latter and more subjective QOL considerations have been secondary to the chemotherapeutic care of the illness. However, there is an increasing emphasis on the inclusion of QOL as a key component in measuring the success of cancer treatment. Given the paucity of QOL data, the expectations of clinical trials is changing in an effort to gain a firmer understanding of the contribution of QOL information to interpreting the results of cancer treatment. In this project, data obtained in a randomized trial of two chemotherapy regimens given as front-line therapy in ovarian cancer were utilized to examine the relationship between chemotherapy-related toxic effects as collected on CRFs (toxicity grades) and the corresponding scored items in a patient-driven QOL assessment. Further assessed was the association between patient-rated global QOL and symptoms as measured by toxicity data collection on CRFs, plus individual QOL questions.

In this trial, the majority of frequently observed treatment-related symptoms were rated as "mild or none" by both the clinicians and the patients. Before chemotherapy, 89% of the symptoms and QOL pairs had a level of agreement greater than 0.80 in the mild or none category. This percentage decreased to 83% during cycles 3 and 6 of the chemotherapy. It was noted that alopecia and QOL question 42 showed only a moderate or fair level of agreement (degree = 0.50 at cycle 3; degree = 0.37 at cycle 6). The patient's interpretation of alopecia in the QOL scale may have differed from the clinicians' rating of alopecia as a toxicity. For example, a clinician may have been influenced by visible hair loss on the head and arms, while a patient may have been less inclined to disclose their hair loss unless specifically questioned by the clinician. However, the self-report form of the QOL scale may have provided the privacy patients desired to disclose the extent of total body hair loss.

In the follow-up phase of the study, after patients were off chemotherapy, 78% of the symptoms-QOL pairs had a high level of agreement (agreement = 0.80) in the mild or none category. Myalgia and QOL question 35, and lethargy and QOL question 18 had slightly lower degrees of agreement (Table 3) at both 3 and 6 months' follow-up. The neurosensory impairment and QOL question 41 had a low

level of agreement at the third and sixth months after patients were off the treatment (degrees of agreement = 0.54 and 0.72, respectively). Further investigation is needed to determine the relevance of using a clinician-rated measure of toxic effects in light of these findings that the neurosensory impairment was the symptom most frequently observed during both the protocol treatment and follow-up. The lack of severe or moderate symptoms, either during chemotherapy or off treatment, makes it difficult to offer any conclusive statements concerning the degree of association between clinically measured toxic effects and the patients' ratings of their own QOL in the severe or moderate category.

The regression analyses performed revealed little difference between the two- and three-item measures for global QOL. During chemotherapy, variables related to gastrointestinal disturbances, urinary frequency, motor impairment, and lethargy remained in the regression equations at both cycles 3 and 6, and so could explain the noted changes in global QOL. In the months following chemotherapy, neurosensory impairment, mood alterations, lethargy, alopecia, and gastrointestinal disturbances (ie, pain and nausea) remained in the regression equations over time, and provided the most information regarding the variability in global QOL assessments. It is interesting to note that urinary incontinence added to the predictor variables at month 3 but was not retained at month 6.

QOL is frequently measured as a secondary outcome in Canadian cancer clinical trials. To determine the best possible measure, a well-established, reliable, and valid tool is critical. Equally important in QOL measurement is the meaningful and relevant dissemination of the QOL data. Clinicians need to be able to identify the key issues that will affect their patients' overall QOL and plan interventions to minimize these distresses.[12] The results from the present analysis supports the contention that the use of the self-report instrument EORTC-QLQ C30+3 and the trial-specific checklist is sensitive to the impact of treatment on QOL for women with ovarian cancer. The matched pairs of QOL items with treatment-related toxic effects collected on CRFs yielded similar findings when the symptoms in question were not severe. EORTC-QLQ C30+3 and the trial-specific checklist captured the impact of the expected treatment effects on QOL, and were sensitive over time. The small number of effects that were described as severe limits conclusions as to the applicability of the self-report questionnaire in assessing such symptoms or toxic effects. However, the available data does not yet rule out the applicability of the approach.

Finally, the present study raises the question of how symptoms and toxicities experienced during treatment should be measured. There seems to be some degree of duplication between the clinician-measured toxicity grades and the patient-rated QOL scores. Methodological issues have remained a concern in obtaining toxicity data.[13,14] This study lends support to the utility of self-report QOL instruments as a useful approach to measurement of cancer patients' experiences of treatment-related symptoms.

Acknowledgment

We thank Eric Bacon, Programmer at the National Cancer Institute of Canada Clinical Trials Group for his continued support in the statistical analysis of this project. Donna Stewart of the Socio-Behavioral Cancer Research Network is also thanked for her contributions in beginning this collaborative project.

Authors' Disclosures of Potential Conflicts of Interest

The authors indicated no potential conflicts of interest.

REFERENCES

1. Canadian Cancer Statistics 2002: Canadian Cancer Society, National Cancer Institute of Canada, Statistics Canada Provincial/Territorial Cancer Registries and Health Canada. http://www.cancer.ca
2. Brundage MD, Davidson J, Mackillop W: Trading treatment toxicity for survival in locally advanced non-small cell lung cancer. J Clin Oncol 15:330-340, 1997
3. Calhoun E, Bennett C, Peeples P, et al: Perceptions of cisplatin-related toxicity among ovarian cancer patients and gynecologic oncologists. Gynecol Oncol 71:369-375, 1998
4. Montazeri A, McEwen J, Gillis C: Quality of life in patients with ovarian cancer: Current state of research. Support Care Cancer 4:169-179, 1996
5. Machin D, Weeden S: Suggestions for the presentation of quality of life data from clinical trials. Stat Med 17:711-724, 1998
6. Piccart M, Bertelsen K, James K, et al: Randomized intergroup trial of cisplatin-paclitaxed versus cisplatin-cyclophosphamide in women with advanced epithelial ovarian cancer: Three-year results. J Natl Cancer Inst 92:699-708, 2000
7. National Cancer Institute of Canada Clinical Trials Group: Expanded Common Toxicity Criteria. Kingston, Ontario, Canada, National Cancer Institute of Canada Clinical Trials Group, 1994, pp 1-35
8. Osoba D, Zee B, Pater J, et al: Psychometric properties and responsiveness of EORTC Quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. Qual Life Res 3:353-364, 1994
9. Feinstein A, Cicchetti D: High agreement but low kappa, I: The problems of two paradoxes. J Clin Epidemiol 43:543-549, 1999
10. Cicchetti D, Feinstein A: High agreement but low kappa, II: Resolving the paradoxes. J Clin Epidemiol 43:551-558, 1990
11. MacKinnon A: A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. Comput Biol Med 30:127-134, 2000
12. Osoba D, Zee B, Pater J, et al: Determinants of postchemotherapy nausea and vomiting in patients with cancer. J Clin Oncol 15:116-123, 1997
13. Savage C, Pater J, Tu D, et al: He said/she said: How much agreement is there on symptoms between common toxicity criteria and quality of life? Proc Am Soc Clin Oncol 21:386a, 2002 (abstr 1540)
14. Brundage M, Pater J, Zee B: Assessing the reliability of two toxicity scales: Implications for interpreting toxicity data. J Natl Cancer Inst 85:1138-1148, 1993